Density of Innovation for Combined Statistical and Metropolitan Areas within the United States

UNDERGRADUATE HONORS CONTRACT RESEARCH THESIS DRAFT

Presented in Partial Fulfillment of the Requirements for the Business Administration Degree with Honors Research Distinction in the Max M. Fisher College of Business at The Ohio State University

By
Samuel Horen

Undergraduate Degree in Business Administration

The Ohio State University
2018

Committee:

Dr. Edward Malecki, Advisor

Dr. Roger Bailey, Advisor

The Ohio State University

Fisher College of Business

2018

ABSTRACT

From a macro perspective, cities across the world are constantly competing to become sticky places by attracting and retaining the best available people and businesses. To remain competitive in this era of rapid technological change and development, city officials and planners have placed a large emphasis on creating innovative environments. This paper hypothesizes that cities with higher population densities are more successful in such attempts, and evaluates whether or not a relationship exists between population density and innovation. To examine this relationship, USPTO patent data from the years 2000-2015 and R&D spending data from the year 2013 are compared against the population density of 26 unique Combined Statistical Areas (CSAs) or Metropolitan Statistical Areas (MSAs). Ultimately this study finds that both patent data and R&D data are weakly positively related to population density. When removing the San Jose-San Francisco-Oakland CSA, a highly innovative outlier, from the dataset the correlation becomes slightly stronger. Finally, this study indicates that population density by CSA may not be the most important population factor to consider when measuring innovation, as total population by CSA provided a stronger correlation to innovation.

ACKNOWLEDGEMENTS

I would like to sincerely thank Dr. Roger Bailey for the guidance and feedback that he has provided throughout the course of this research project.

I would also like to sincerely thank Dr. Edward Malecki for providing direction through his vast technical expertise and insight on the subject matter.

VITA

June 2014 --Farmington High School

May 2018---BSBA Finance, The Ohio State University

LIST OF TABLES

LIST OF FIGURES

Table of Contents

I. INTRODUCTION:

Innovation is a key component of growth and development in the global economic system. It is an important catalyst that drives new businesses and improves currently established firms. In modern times, there have been great efforts made amongst geographers and economists to quantify and collect data regarding innovation. However, this data can be very difficult to collect, as it is oftentimes deemed confidential by private firms that do not want to lose their competitive edge. As a result several metrics have become the preferred way to measure innovation. Examining the number of patents published is the standard way to measure innovative output. Conversely, the amount of research and development spending that has occurred is a common way to measure innovative input. This research paper will seek to use these variables to examine the relationship between innovation and population density.

In the past, various studies have explored the correlation between total population and innovation levels. However, total population is oftentimes a misleading way to measure the agglomeration of an area. For example, according to the latest US census data, the combined statistical area of Denver has roughly the same population (3,470,235) as the combined statistical area of Cleveland (3,483,311). However, Denver contains over two times more land than Cleveland (33,824 km^2 vs. 15,231 km^2). As a result the urban landscape of Cleveland and Denver paint two very different pictures. For this reason, population density may be a more important variable to take into consideration, as it offers a more holistic and fair measure of a city's agglomeration and organization.

In previous total population vs. innovation studies there have been mixed results. Older research has indicated that innovation is moving away from cities towards suburban office parks, however newer studies seem to show that this trend could be reversing. As a result it is expected that a positive correlation between population density and innovation will be found, since total population and population density have some commonalities. Finally, as a disclaimer, since population density will be used and innovation will be tracked with a combination of both patents and research and development spending this study may yield different results than in the past.

Overall this research is very important for several groups of people. From a city governmental point of view, these results might help lead zoning decision-making, as there may be an optimal size/organization of agglomeration that encourages innovation and economic growth. From a business point of view, this study may help indicate where to locate in order to increase innovative potential.

The following sections will walk through some of the finer details of this research idea. First the literature review will delve further into the current research that has already been done in this area. The various studies that have addressed similar research questions will be discussed, and ultimately this section will attempt to show how this paper will build off these ideas and conclusions. Next, in the hypothesis section, the important terms for this study will be precisely defined and formalized, and expectations about results will be discussed. The methodology section, which follows, will discuss how patent and R&D spending data will be utilized within this paper. The results and findings section will be dedicated to examining the raw data and trends that exist. Finally the discussion and conclusion section will tie together the outcomes discovered in the previous section and will suggest further ways that innovation can be researched and explored as a result of this paper.

II. LITERATURE REVIEW:

There have been various research studies conducted that compare population to innovation. However, population density is rarely referenced and innovation is often measured through only one lens.

The most important piece of literature regarding innovation research is the Oslo Manual (OECD, 2005). This manual is expansive, and has a sizeable amount of pertinent information. Essentially, it examines the definition and measurement of innovation, a seemingly abstract term. At its core, the manual breaks innovation down into four areas: product, process, marketing, and organizational innovations. From there, it offers various ways to measure innovation from each of the specific areas. To best obtain data for process, marketing, and organizational innovation, various extensive surveys would have to be conducted. However, for product innovation the manual suggests that utilizing Research and Development expenditure to measure general innovative input and utilizing patent data to measure innovative output is most effective.

(Packalen and Bhattacharya, 2015) offered a history of US patents over the time period of 1836-2010. Ultimately, this paper provides evidence that innovation appears to be moving away from larger cities. Although this paper measures population density vs. patent innovation it does not include research and development in its analysis and has data that is not as current as the data that is utilized in this paper. Since the most recent recession, various economic and geographic trends have changed at a rapid pace and are worth reexamining.

(Forman, Goldfarb, and Greenstein 2016) is an example of another paper that uses patents to measure innovation. The results and data show that patents in the San Francisco bay area continue to increase at a rate much higher than elsewhere in the US. Additionally, this paper did an excellent job of framing patents as a useful measure of invention.

R&D is analyzed as a measure of innovation in (Crescenzi and Rodríguez-Pose 2013). This paper ultimately concludes that local R&D spend is a predictive measure of innovative performance. As a result, it is important to include R&D in innovative analysis.

Another example of population density being looked at specifically is in (Florida, 2016). This study analyzed the flow of venture capital vs. population density. Data was analyzed briefly at a city level, then more specifically at a local level by looking at effects on a zip code basis for New York City, Los Angeles, San Francisco, and Boston. In addition, various other geographic factors were examined vs. innovation, such as poverty level and housing costs. Florida showed that venture capital is most prevalent in dense regions, which might suggest a shift in innovation from suburbs and edge cities back towards densely populated urban regions. Venture capital represents money being invested into an idea an investor likes and thinks could develop into something larger. It is therefore more of a forward looking number and is not necessarily an accurate indication of the current state of innovation.

(Duranton and Puga 2001) and (OhUallicháin, 2000) help explain why innovation is such an important metric to measure. These papers show how innovation drives growth and development within a city, and show how various factors influence innovation. Specifically, total population, diversity, and specialization are examined. These three factors appear to greatly influence innovation within cities, and work best when combined. The data from the former study comes from French urban regions, and the resulting analysis is difficult to interpret and not exactly relatable, however it provides great insight into the drivers of innovation. The results from the latter study is conversely very easy to interpret and shows that throughout the 1990's population has been moderately correlated to the level of patents produced.

Just this past summer a new paper (Berkes and Gaetani 2017) emerged that directly studied the role of population density as a driver of innovation. This study georeferenced patent information provided by the USPTO from the years 2002-2014 at the County Sub-Division (CSD) level. They found that density plays a smaller than expected role in generating innovation, as over 40% of patents originate from counties with a population below 1,000 people per square kilometer. In this paper Berkes and Gaetani also provide reasons as to why dense spaces generate more innovative results. Their primary conclusion is that population density allows for a higher level of diverse learning opportunities. In more dense areas people are exposed to individuals with different backgrounds more frequently and as result these interactions spark more creativity and learning.

Another important paper was published by the Brookings Institute in 2017 (Shambaugh, Nunn, Portman 2017). This paper has 11 major conclusions relating to patents, R&D spending, and ultimately the relationship between this data and innovative activity. One of the most important takeaways is that the US patent office currently receives six times as many applications as it did in 1980 suggesting that older data may not be relevant in evaluating modern innovative trends. Additionally this study found R&D investment leads to the discovery of new technologies, but often only after many years. This validates the importance of considering R&D when measuring innovation. Lastly this study found that 59 percent of patents were awarded to applicants in 20 metropolitan areas within the US representing only 36 percent of the population. This statistic suggests that innovation is concentrated in dense metropolitan areas.

Lastly, (Crescenzi, Rodriguez-Pose, and Storper 2007) shines light on other variables that influence innovation by comparing innovation in Europe and in the United States. In this analysis it is apparent that higher mobility of capital, population, and knowledge allows the US to exploit local synergies and as a result drive higher levels of innovation.

Overall, these studies are applicable and closely related to the research question examined in this paper. However, none of these papers directly utilized Combined Statistical Area population and land area as a measure of population density which is a central part of this paper. Combined Statistical Areas include the greater urban sprawl of a large city or densely populated region. In addition no papers included research and development data to measure innovation, and some of these papers referenced data collected prior to or during the time of the last recession. Since a lot has happened economically and geographically in the last 10 years, it is useful to find new data and look for current trends post-recession.

Hypothesis

As will be discussed more thoroughly in the methodology section, this study will ultimately be examining the interaction between two key variables: population density and innovation.

Since studies mentioned in the literature review have varying results regarding the measured relationship between total population and innovation, new data might yield different conclusions. It is therefore predicted that withholding significant outliers, population density and innovation will be positively correlated. This hypothesis is based off the findings discovered by Berkes and Gaetani which suggest that the diverse social interactions enabled by population density, can potentially drive innovation. However it is also expected that this trend might not be linear forever, as overcrowding can lead to counter productivity. Berkes and Gaetani also showed that within dense regions higher population density does not necessarily yield more innovation.

Additionally, as Florida showed there are many variables that influence innovation, and even if a strong correlation exists, causation will not be proven. As a result, this paper will recognize that innovation is very complex, and as a result is offering a view into just one of the potential driving forces.

III. METHODOLGY:

In this section the definitions for population density and innovation will be explored thoroughly, and general set up of the data collection/analysis will be discussed.

To measure population density, the population of a specific Combined Statistical Area (CSA) or Metropolitan Statistical Area (MSA) will be divided by the land area, per square kilometer, that encompasses the area's population. This data is accessible through the US Census Bureau.

The other variable being investigated is innovation, which is conversely extremely difficult to define and measure. To define innovation, the Oslo manual will be referenced. In essence the manual suggested that innovation can be broken down into four categories; product, process, marketing, and organizational innovations. In this context, innovation simply means an incremental improvement over the status quo, whether that means a new product, process, or idea is generated, or a current product, process, or idea is improved. The manual clearly indicates that the latter three, process, marketing, and organizational innovation, are practically impossible to measure accurately alone, and would require significant resources and time to individually reach out to all firms and objectively decide and quantify this type of innovation. Product innovation, the first type of innovation referenced in the manual, is more hands on and has been measured before by looking at patents. While patents indicate the quantity of innovation occurring in a particular area, they do not necessarily relate to quality. Just because a patent is granted, does not necessarily mean the product or idea will be successful or create tremendous change. As a result utilizing patent information to measure innovation is helpful but by no means absolute.

Another way the Oslo Manual recommended measuring innovation is by looking at Research and Development spending data. While this is a direct measure of the inputs that go into innovation, it is an additional indicator of the quantity of innovation that can be expected to occur in a specific region looking forward. R&D offers a more holistic view of innovation as it tends to include more process, marketing, and organizational innovation than patents encompass. While both of these measures of innovation are not perfect, they are quantifiable and can provide a glimpse into the innovation occurring. The data for patents can be found through the USPTO and the R&D data can be found through nsf.gov, which has collected the amount of R&D spending in various CSAs and MSAs. Additionally, SSTI is a website that outlines where to find datasets for various geographic measures such as patents and R & D spending.

To analyze this data and reach conclusions, regressions will be run on the data. Three major comparisons will be examined: CSA population density vs. CSA patents, CSA population density vs. CSA R&D spending, and CSA Patents vs. CSA R&D spending. Furthermore, additional regressions will be run which do not include the San Jose-San Francisco-Oakland CSA. For the purposes of this study this CSA can be considered an outlier since it includes Silicon Valley, which includes a disproportionately high amount of innovation. While the level of population density for this CSA may be ideal for innovation to occur, it is worthwhile to explore the data without this data point since it skews the results and correlation levels. Additionally, it is worth noting that other variables such as the name and brand of this region has drawn a lot of entrepreneurs and inventors, which has therefore driven up the quantity of innovation data.

Unfortunately, due to the scarcity and difficulty of collecting R&D data, the results are limited to the 26 CSAs or MSAs that appear in the NSF dataset.

IV. RESULTS AND FINDINGS

Refined Data:

In order to analyze the effect that population density has on innovation it was first important to find and then refine the proper data discussed in the methodology section. The first table represents the R&D spend data by 26 of the largest CSAs and MSAs.

TABLE 1. Business R&D performed and paid for by large-R&D companies at their largest R&D location, by geographic area: 2013

Combined statistical area or metropolitan statistical area	Companies reporting largest R&D location (number)	R&D performance at largest location ($millions)
All locations	2,946	123,278
Atlanta-Athens-Clarke County-Sandy Springs, GA CSA	45	1,020
Austin-Round Rock, TX MSA	33	860
Boston-Worcester-Providence, MA-RI-NH CSA	240	6,108
Chicago-Naperville, IL-IN-WI CSA	115	2,297
Cleveland-Akron-Canton, OH CSA	36	676
Dallas-Fort Worth, TX CSA	44	1,291
Denver-Aurora, CO CSA	54	630
Detroit-Warren-Ann Arbor, MI CSA	93	9,382
Grand Rapids-Wyoming-Muskegon, MI CSA	26	310
Hartford-West Hartford, CT CSA	22	1,426
Houston-The Woodlands, TX CSA	60	1,972
Los Angeles-Long Beach, CA CSA	197	13,505
Miami-Fort Lauderdale-Port St. Lucie, FL CSA	22	684
Milwaukee-Racine-Waukesha, WI CSA	37	679
Minneapolis-St. Paul, MN-WI CSA	90	2,045
New York-Newark, NY-NJ-CT-PA CSA	209	7,913
Philadelphia-Reading-Camden, PA-NJ-DE-MD CSA	73	2,447
Phoenix-Mesa-Scottsdale, AZ MSA	31	1,038
Pittsburgh-New Castle-Weirton, PA-OH-WV CSA	38	462
Portland-Vancouver-Salem, OR-WA MSA	37	3,895
Raleigh-Durham-Chapel Hill, NC CSA	26	984
Salt Lake City-Provo-Orem, UT CSA	49	573
San Diego-Carlsbad, CA MSA	96	4,762
San Jose-San Francisco-Oakland, CA CSA	381	29,547
Seattle-Tacoma-Olympia, WA CSA	72	10,522
Washington-Baltimore-Arlington, DC-MD-VA-WV CSA	79	1,715
All other geographic areas reported as largest location	741	16,534

CSA = combined statistical area; MSA = metropolitan statistical area.

NOTES: R&D data are for companies known to have performed at least $3 million of R&D in prior years that reported their largest R&D location. Only geographic areas where at least 20 companies report their largest location are listed.

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, and U.S. Census Bureau, Business R&D and Innovation Survey, 2013.

The table above explicitly displays the number of large R&D companies reporting their R&D spending, and the amount of R&D spending in millions sorted at a CSA/MSA level. One potential source of error that may be present in this dataset, is that cost of living is not taken into account. Certain MSA's and CSA's have a significantly higher cost of living and as a result may yield higher values for R&D, since the salaries of the people performing the R&D will be higher. However, this data undoubtedly provides insight into where large companies are setting up and investing in innovative input. While this spending and innovative input does not guarantee innovation, it ultimately is one of the largest sources and origins of innovation.

The next table provides patent information for the same 26 CSAs and MSAs provided for the R&D data.

TABLE 2. Patent count for business in selected CSA and MSA (2000-2015)

Combined statistical area or metropolitan statistical area	Patent Count
Metropolitan	1,518,075
Micropolitan	54,038
Non Metro/Micropolitan Statistical Area	20,013
Undetermined Statistical Area	40
All Locations	1,592,166
Atlanta-Athens-Clarke County-Sandy Springs, GA CSA	23,586
Austin-Round Rock, TX MSA	33,753
Boston-Worcester-Providence, MA-RI-NH CSA	76,798
Chicago-Naperville, IL-IN-WI CSA	46,991
Cleveland-Akron-Canton, OH CSA	18,726
Dallas-Fort Worth, TX CSA	34,898
Denver-Aurora, CO CSA	11,436
Detroit-Warren-Ann Arbor, MI CSA	45,270
Grand Rapids-Wyoming-Muskegon, MI CSA	3,006
Hartford-West Hartford, CT CSA	8,920
Houston-The Woodlands, TX CSA	32,197
Los Angeles-Long Beach, CA CSA	74,381
Miami-Fort Lauderdale-Port St. Lucie, FL CSA	17,019
Milwaukee-Racine-Waukesha, WI CSA	10,315
Minneapolis-St. Paul, MN-WI CSA	41,696
New York-Newark, NY-NJ-CT-PA CSA	92,577
Philadelphia-Reading-Camden, PA-NJ-DE-MD CSA	33,573
Phoenix-Mesa-Scottsdale, AZ MSA	20,933
Pittsburgh-New Castle-Weirton, PA-OH-WV CSA	11,117
Portland-Vancouver-Salem, OR-WA MSA	26,331
Raleigh-Durham-Chapel Hill, NC CSA	22,811
Salt Lake City-Provo-Orem, UT CSA	10,410
San Diego-Carlsbad, CA MSA	45,465
San Jose-San Francisco-Oakland, CA CSA	233,454
Seattle-Tacoma-Olympia, WA CSA	44,924
Washington-Baltimore-Arlington, DC-MD-VA-WV CSA	35,334
All other geographic areas reported as largest location	536,245

CSA = combined statistical area; MSA = metropolitan statistical area.

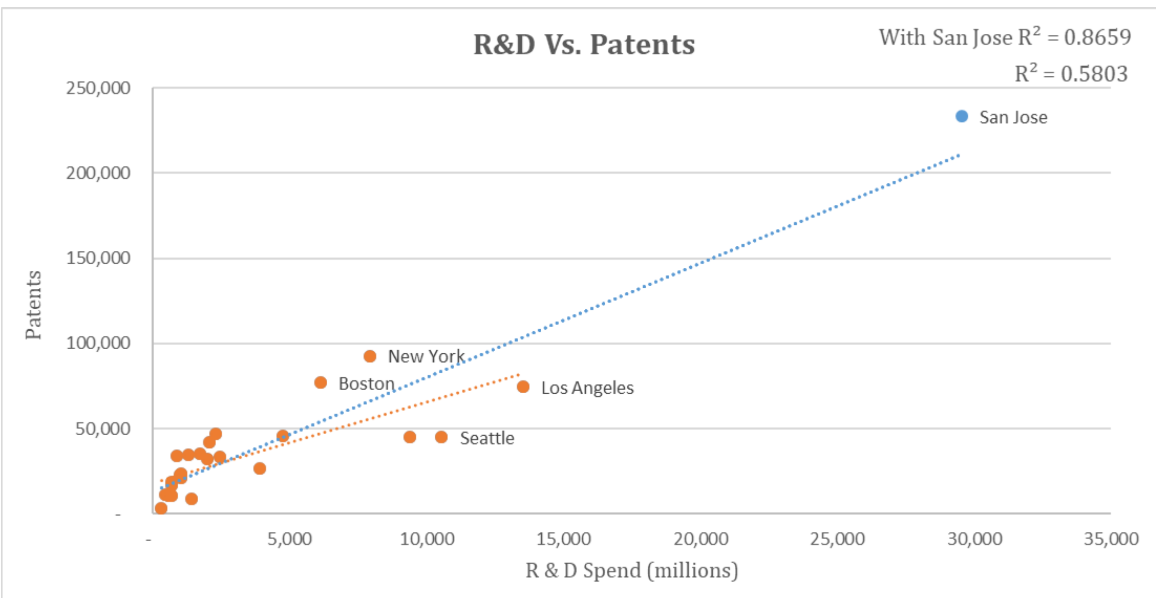
NOTES: Defines metropolitan and micropolitan statistical areas according to published standards that are applied to Census Bureau data. The general concept of a metropolitan or micropolitan statistical area is that of a core area containing a substantial population nucleus, together with adjacent communities having a high degree of economic and social integration with that core. The data found in this table, is prepared from the Technology Assessment and Forecast (TAF) database, which profiles utility patents (i.e., ' patents for inventions ') granted during the time period specified in the report.

SOURCE: The patent data used to prepare this report were derived from the USPTO's Technology Assessment and Forecast database

Although patent count may seem like an elementary and plain way to measure innovation, it remains one of the best quantitative ways to do so with the limited public data that exists on the topic. The patent information provided is for all patents types which cover any product innovation filed through the USPTO.

To show the relationship between R&D (the input of innovation) and patents (the output of innovation) the two variables are graphed against each other for the 26 CSAs or MSAs being examined.

GRAPH 1



With an R^2 of roughly .87 (with San Jose) and an R^2 of .5803 (without San Jose) this graph shows that there is a fairly strong correlation between R&D spend and patents. This indicates that R&D is a fair way to predict the current and potential futuristic state of innovation for the selected CSAs and MSAs. Since R&D spending is not perfectly correlated to patents produced, it is therefore important to analyze the variables separately. However, this relationship provides validity for the connection between population density and R&D spending, and demonstrates why it is a variable that is worth considering when taking innovation into account.

The subsequent table is the condensed version of population and land area (km^2) for the specified CSAs and MSAs for this study.

TABLE 3. Population and density data for selected CSA and MSA (2016 census estimate)

Combined statistical area or metropolitan statistical area	Land Area (km²)	Population (2016 census estimate)	Population Density (population / km²)
Atlanta-Athens-Clarke County-Sandy Springs, GA CSA	30,672	6,451,262	210
Austin-Round Rock, TX MSA	10,933	2,056,405	188
Boston-Worcester-Providence, MA-RI-NH CSA	25,129	8,176,376	325
Chicago-Naperville, IL-IN-WI CSA	27,548	9,882,634	359
Cleveland-Akron-Canton, OH CSA	15,231	3,483,311	229
Dallas-Fort Worth, TX CSA	42,697	7,673,305	180
Denver-Aurora, CO CSA	33,824	3,470,235	103
Detroit-Warren-Ann Arbor, MI CSA	16,915	5,318,653	314
Grand Rapids-Wyoming-Muskegon, MI CSA	13,264	1,443,508	109
Hartford-West Hartford, CT CSA	5,646	1,476,637	262
Houston-The Woodlands, TX CSA	32,431	6,972,374	215
Los Angeles-Long Beach, CA CSA	87,944	18,688,022	212
Miami-Fort Lauderdale-Port St. Lucie, FL CSA	19,328	6,723,472	348
Milwaukee-Racine-Waukesha, WI CSA	9,776	2,043,274	209
Minneapolis-St. Paul, MN-WI CSA	28,834	3,894,820	135
New York-Newark, NY-NJ-CT-PA CSA	35,880	23,689,255	660
Philadelphia-Reading-Camden, PA-NJ-DE-MD CSA	19,000	7,179,357	378
Phoenix-Mesa-Scottsdale, AZ MSA	37,724	4,661,537	124
Pittsburgh-New Castle-Weirton, PA-OH-WV CSA	18,251	2,635,228	144
Portland-Vancouver-Salem, OR-WA MSA	32,926	3,160,488	96
Raleigh-Durham-Chapel Hill, NC CSA	14,273	2,156,253	151
Salt Lake City-Provo-Orem, UT CSA	60,494	2,514,748	42
San Diego-Carlsbad, CA MSA	10,895	3,317,749	305
San Jose-San Francisco-Oakland, CA CSA	26,260	8,751,807	333
Seattle-Tacoma-Olympia, WA CSA	31,834	4,684,516	147
Washington-Baltimore-Arlington, DC-MD-VA-WV CSA	32,709	9,665,892	296

CSA = combined statistical area; MSA = metropolitan statistical area.

SOURCE: US Census Bureau

The data above is generated from the US Census Bureau and estimates the 2016 population and land area for the specific CSA/MSA being referenced. To calculate the third column population was divided by land area. This produced a number which represents the estimated density of population per square kilometer for the 26 observed CSAs and MSAs.

Before delving into the heart of the results sections, one other important thing to note is that certain sectors historically produce more patents than other sectors. Given the nature of this inequity of distribution for R&D spend and patents across all sectors, it is expected that the data in this paper is not necessarily painting a holistic picture of innovation.

Overall View of Innovation

%	Industry	2014 Volume	2013 Volume	% Change
5%	Aerospace & Defense	62,162	63,080	-1%
12%	Automotive	153,872	152,221	1%
3%	Biotechnology	42,584	39,685	7%
1%	Cosmetics & Well Being	11,017	10,197	8%
2%	Food, Tobacco & Beverage Fermentation	26,333	21,758	21%
6%	Home Appliances	71,278	71,118	0%
30%	Information Technology	380,325	367,028	4%
7%	Medical Devices	93,462	99,290	-6%
2%	Oil & Gas	24,158	23,925	1%
9%	Pharmaceuticals	111,479	99,950	12%
9%	Semiconductors	112,625	119,099	-5%
13%	Telecommunications	161,739	153,153	6%

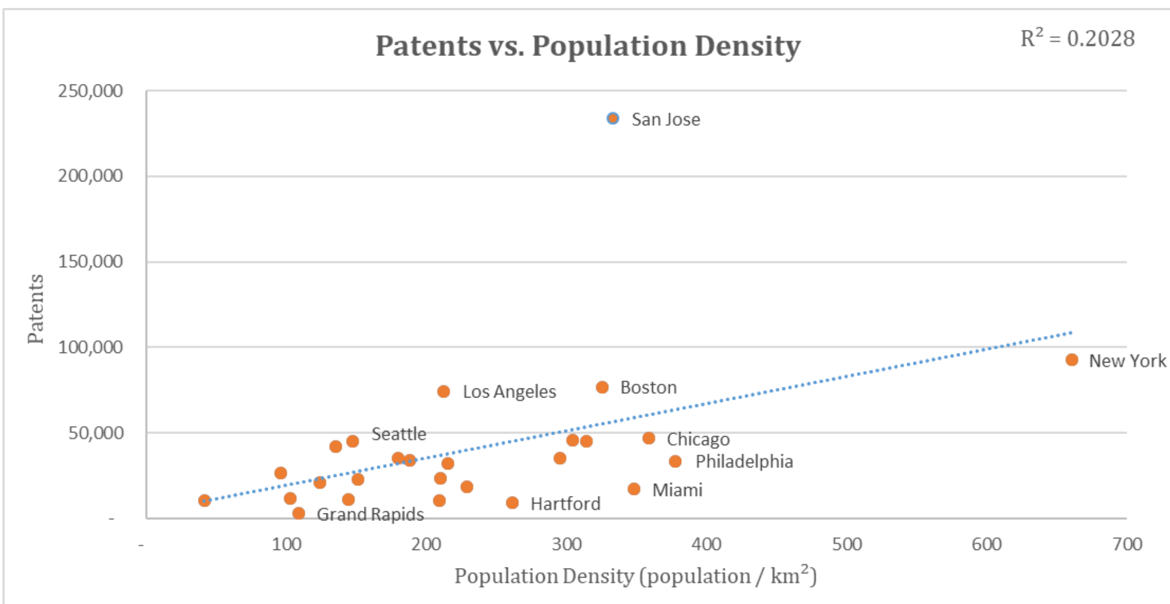
Source: Thomson Reuters Derwent World Patents Index

By viewing the above report provided by the Thompson Reuters Derwent World Patents Index it is rather clear that patent data has a strong bias towards the information technology, automotive, and telecommunication sectors which combined produce more than 55% of total 2014 patents alone.

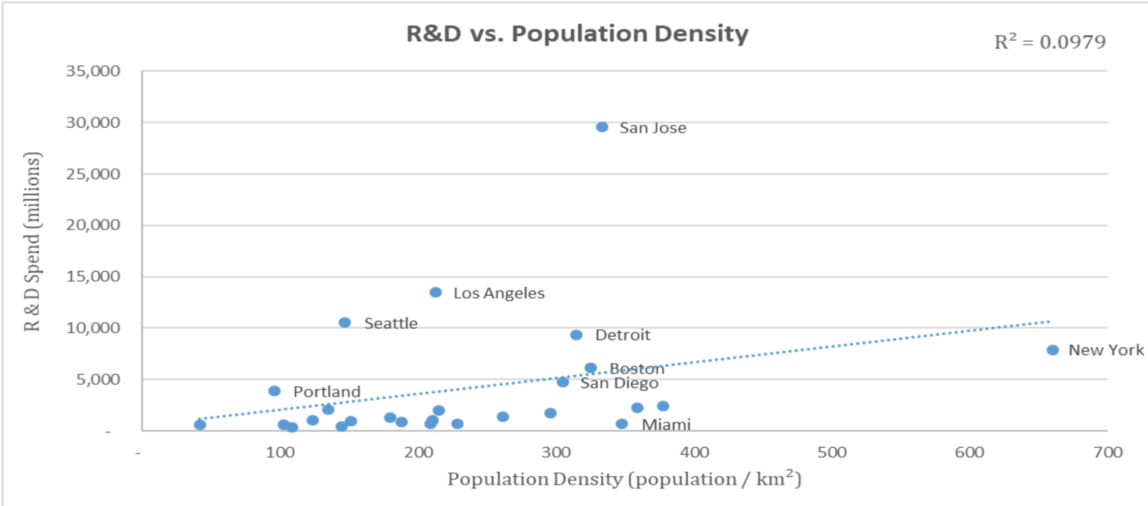
Regression Findings:

For the first two regressions the patent data by CSA/MSA and R&D spending data by CSA/MSA were compared to the corresponding CSA/MSA population density. The results were as follows:

GRAPH 2



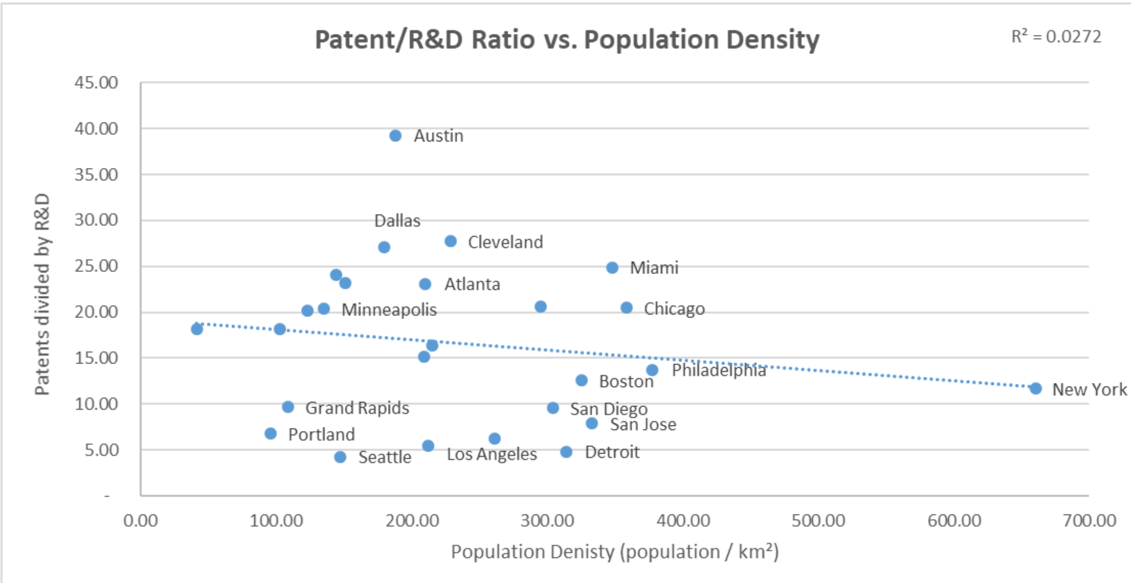
GRAPH 3



With an R^2 of .2028 for the patent data and an R^2 of .0979 for R&D data these graphs seem to indicate that although population density is positively correlated to innovative inputs and outputs, the relationship is quite weak. To further explore this relationship the San Jose CSA was removed and the regression were run again. This was done since San Jose has a disproportionately high amount of innovative activity due to the presence of Silicon Valley.

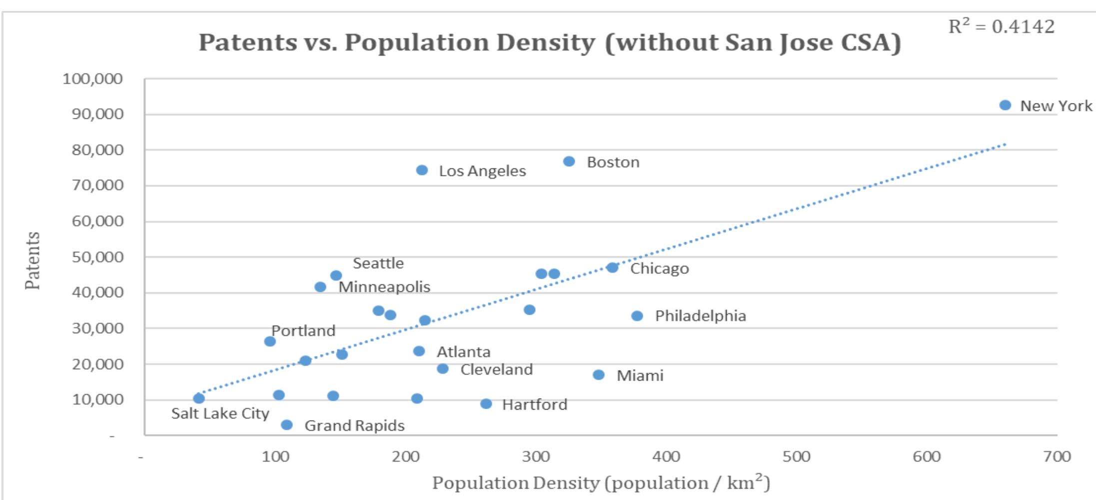
To further document the relationship between the innovation and population density, Graph 4 displays the correlation between patents divided by R&D vs. population density.

GRAPH 4

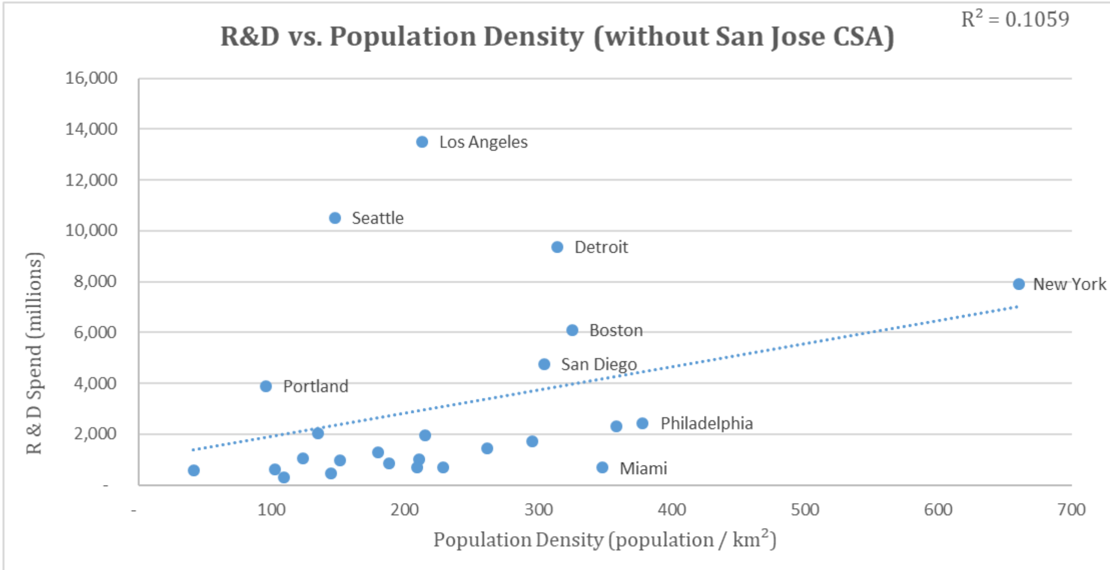


This relationship which is nearly nonexistent as evidenced by the low R^2 value, shows that perhaps the hypothesized relationship is less accurate than the previous graphs indicate. Perhaps this is because of reverse causality, or a discrepancy between the effectiveness of cities converting R&D into patents.

GRAPH 5



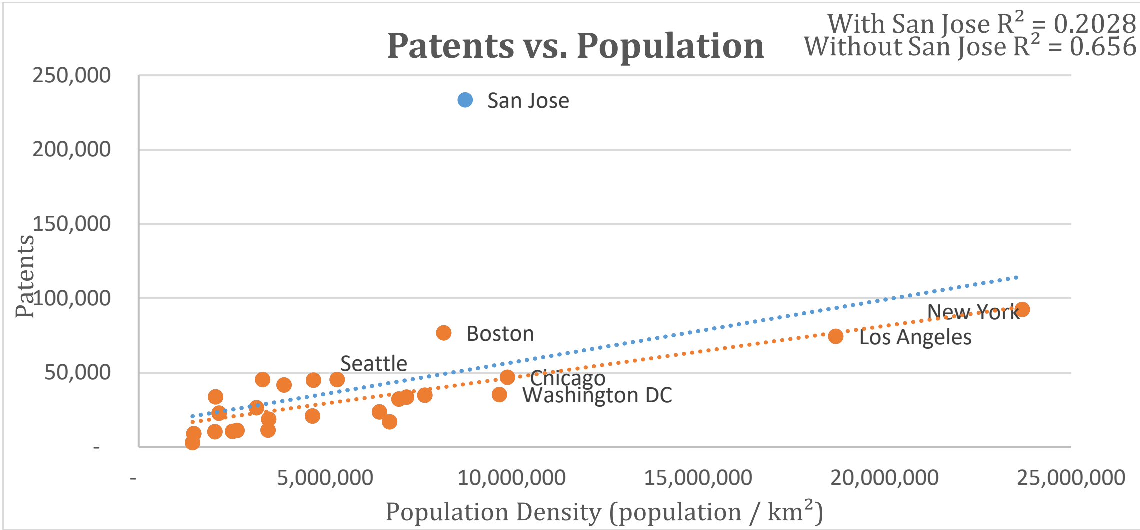
GRAPH 6



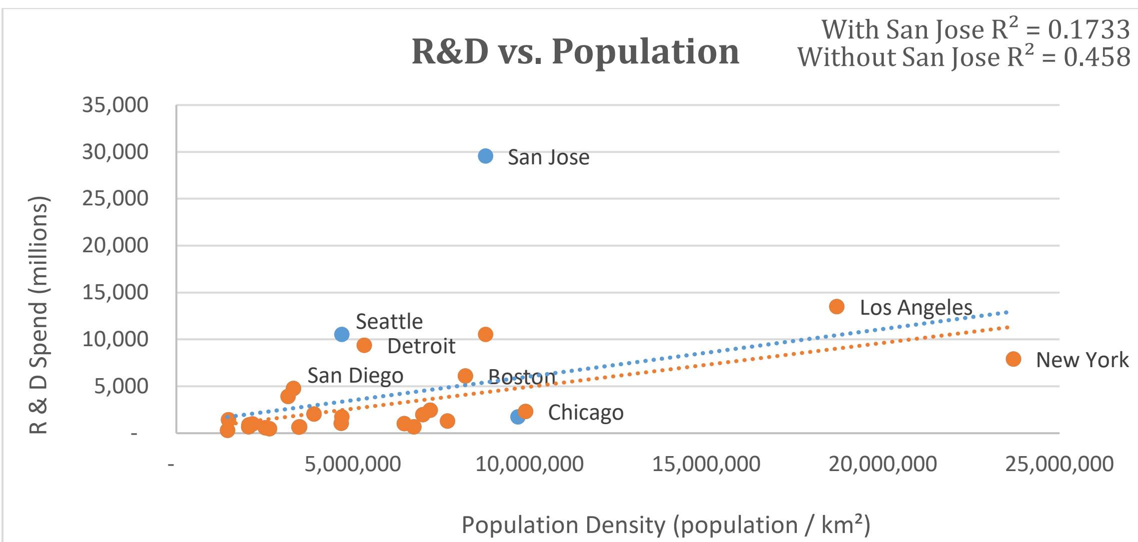
As evidenced by the graphs above the patent R^2 jumped up to .4142 from .2028 while the R&D spend R^2 remained fairly stagnant at .1059 slightly up from .0979. Although, San Jose is an innovation hub it only seemed to skew the patent vs. population density regression.

To further analyze the effect that population size has on innovation, patent and R&D data was plotted vs. total population in the 26 CSAs and MSAs.

GRAPH 7



GRAPH 8



This yielded very similar results to the regressions provided by the population data statistics. The R^2 for patents moved from .2028 to .234 and the R^2 for R&D moved from .0979 to .1733 with San Jose. However without San Jose the R^2 values increased dramatically for both .656 and .458 respectively. This is supported by other historic papers and further indicates that an increase in population tends to lead to an increase in innovation.

ANALYSIS

There are a few key takeaways from the regressions found in the results and findings section as well as this paper as a whole. First, population density within CSAs and MSAs, while positively related to the number of patents produced and R&D spend occurring, is not strongly correlated to innovation. This finding is congruous with Berkes and Gaetani (2017) and indicates that perhaps the disadvantages of population density outweigh its advantages at a certain point. Second, CSA's and MSA's are perhaps too large of an area to utilize for innovation measurement purposes. Some of these regions span thousands of km^2 and reach across various states and regions. In these large areas, the space is undoubtedly heterogeneously populated and as a result one sweeping population density number does not accurately depict the organization of the agglomeration. Third, this paper does not indicate where the innovation is occurring at a micro level within the CSA or MSA. For example, is the activity occurring in the denser parts of the city or in the suburban office parks?

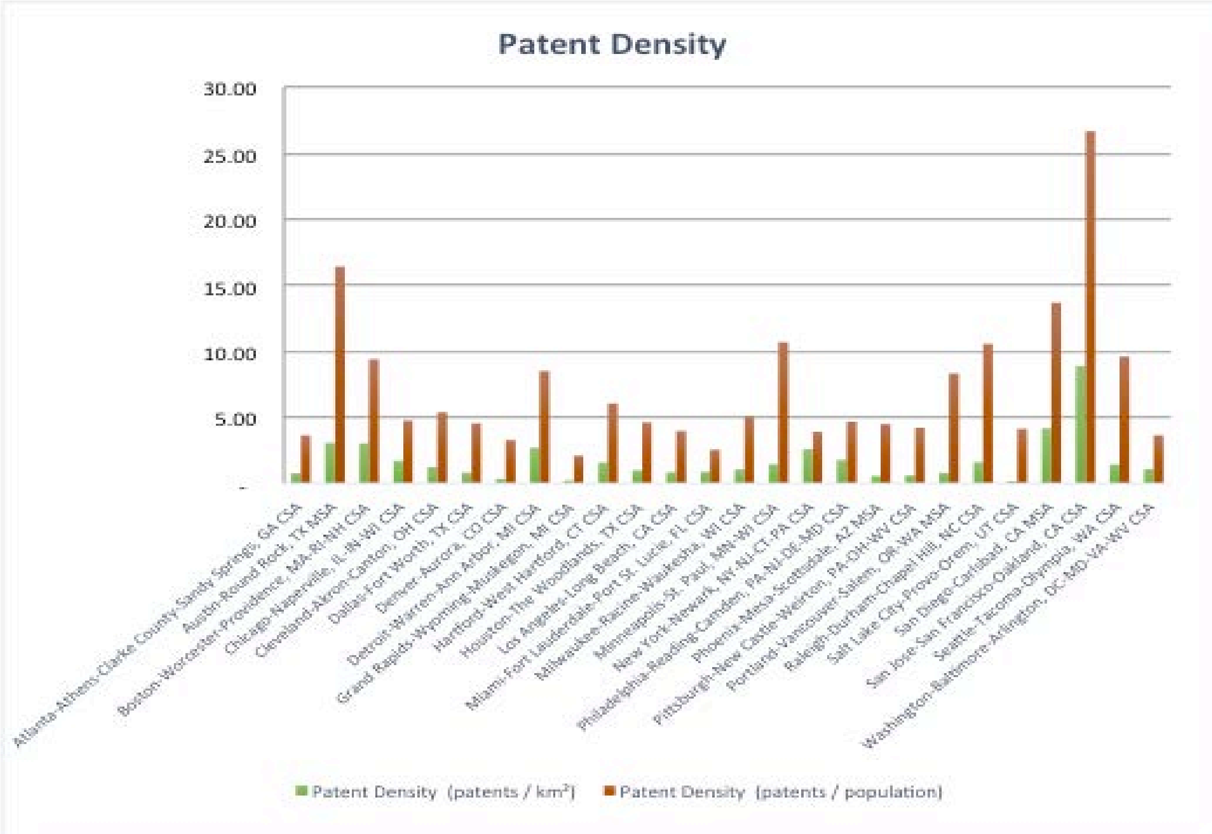
The following table and graphs further examine the primary observed conclusions. First table 4 takes the data in tables 1-3 and produces a few helpful metrics to better conceptualize the innovation that is occurring.

TABLE 4. Patent and R&D density analysis

	Patent Density (patents / km²)	R&D Density (R&D spend / km²)	Patent Density (patents / population)	R&D Density (R&D spend / population)	Patent Density (patents / population density)	R&D Density (R&D spend / population density)
Atlanta-Athens-Clarke County-Sandy Springs, GA CSA	0.77	0.03	3.66	0.16	112.14	4.85
Austin-Round Rock, TX MSA	3.09	0.08	16.41	0.42	179.45	4.57
Boston-Worcester-Providence, MA-RI-NH CSA	3.06	0.24	9.39	0.75	236.03	18.77
Chicago-Naperville, IL-IN-WI CSA	1.71	0.08	4.75	0.23	130.99	6.40
Cleveland-Akron-Canton, OH CSA	1.23	0.04	5.38	0.19	81.88	2.96
Dallas-Fort Worth, TX CSA	0.82	0.03	4.55	0.17	194.18	7.18
Denver-Aurora, CO CSA	0.34	0.02	3.30	0.18	111.47	6.14
Detroit-Warren-Ann Arbor, MI CSA	2.68	0.55	8.51	1.76	143.97	29.84
Grand Rapids-Wyoming-Muskegon, MI CSA	0.23	0.02	2.08	0.21	27.62	2.85
Hartford-West Hartford, CT CSA	1.58	0.25	6.04	0.97	34.11	5.45
Houston-The Woodlands, TX CSA	0.99	0.06	4.62	0.28	149.76	9.17
Los Angeles-Long Beach, CA CSA	0.85	0.15	3.98	0.72	350.03	63.55
Miami-Fort Lauderdale-Port St. Lucie, FL CSA	0.88	0.04	2.53	0.10	48.92	1.97
Milwaukee-Racine-Waukesha, WI CSA	1.06	0.07	5.05	0.33	49.35	3.25
Minneapolis-St. Paul, MN-WI CSA	1.45	0.07	10.71	0.53	308.68	15.14
New York-Newark, NY-NJ-CT-PA CSA	2.58	0.22	3.91	0.33	140.22	11.99
Philadelphia-Reading-Camden, PA-NJ-DE-MD CSA	1.77	0.13	4.68	0.34	88.85	6.48
Phoenix-Mesa-Scottsdale, AZ MSA	0.55	0.03	4.49	0.22	169.40	8.40
Pittsburgh-New Castle-Weirton, PA-OH-WV CSA	0.61	0.03	4.22	0.18	76.99	3.20
Portland-Vancouver-Salem, OR-WA MSA	0.80	0.12	8.33	1.23	274.32	40.58
Raleigh-Durham-Chapel Hill, NC CSA	1.60	0.07	10.58	0.46	150.99	6.51
Salt Lake City-Provo-Orem, UT CSA	0.17	0.01	4.14	0.23	250.42	13.78
San Diego-Carlsbad, CA MSA	4.17	0.44	13.70	1.44	149.30	15.64
San Jose-San Francisco-Oakland, CA CSA	8.89	1.13	26.67	3.38	700.48	88.66
Seattle-Tacoma-Olympia, WA CSA	1.41	0.33	9.59	2.25	305.28	71.50
Washington-Baltimore-Arlington, DC-MD-VA-WV CSA	1.08	0.05	3.66	0.18	119.57	5.80

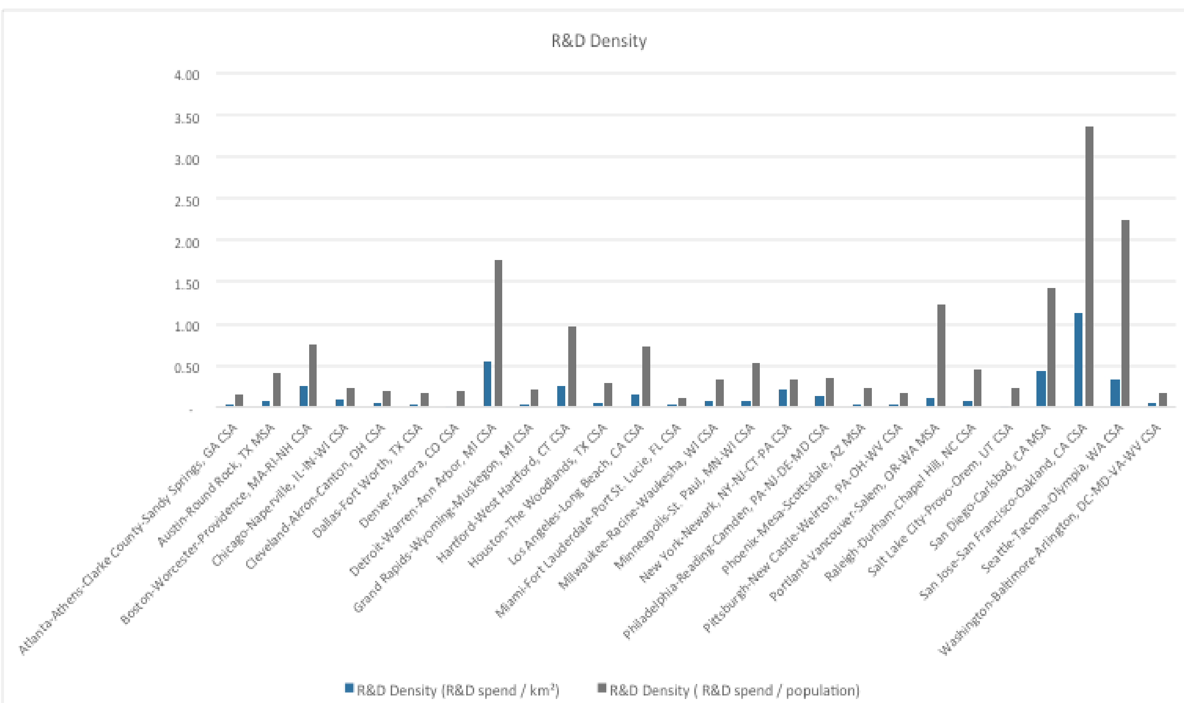
In columns 7 and 8 patent density and R&D spend density is determined by dividing the number of respective units by the CSA/MSA land area. In columns 9 and 10 the respective patent and R&D data is divided by the corresponding CSA/MSA population. These results are combined and outlined below in graphs 9 and 10. Graph 9 looks solely at the data provided by patents, and graph 10 looks solely at the data provided by R&D spend.

GRAPH 9



Patent Density
30.00
25.00
20.00
15.00
10.00
5.00
Atlanta-Athens-Clarke County-Sandy Springs, GA CSA
Austin-Round Rock, TX MSA
Boston-Worcester-Providence, MA-RI-NH CSA
Chicago-Naperville, IL-IN-WI CSA
Cleveland-Akron-Canton, OH CSA
Dallas-Fort Worth, TX CSA
Denver-Aurora, CO CSA
Detroit-Warren-Ann Arbor, MI CSA
Grand Rapids-Wyoming-Muskegon, MI CSA
Hartford-West Hartford, CT CSA
Houston-The Woodlands, TX CSA
Los Angeles-Long Beach, CA CSA
Miami-Fort Lauderdale-Port St. Lucie, FL CSA
Milwaukee-Racine-Waukesha, WI CSA
Minneapolis-St. Paul, MN-WI CSA
New York-Newark, NY-NJ-CT-PA CSA
Philadelphia-Reading-Camden, PA-NJ-DE-MD CSA
Phoenix-Mesa-Scottsdale, AZ MSA
Pittsburgh-New Castle-Weirton, PA-OH-WV CSA
Portland-Vancouver-Salem, OR-WA MSA
Raleigh-Durham-Chapel Hill, NC CSA
Salt Lake City-Provo-Orem, UT CSA
San Diego-Carlsbad, CA MSA
San Jose-San Francisco-Oakland, CA CSA
Seattle-Tacoma-Olympia, WA CSA
Washington-Baltimore-Arlington, DC-MD-VA-WV CSA
Patent Density (patents / km²)
Patent Density (patents / population)

GRAPH 10



R&D Density
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
Atlanta-Athens-Clarke County-Sandy Springs, GA CSA
Austin-Round Rock, TX MSA
Boston-Worcester-Providence, MA-RI-NH CSA
Chicago-Naperville, IL-IN-WI CSA
Cleveland-Akron-Canton, OH CSA
Dallas-Fort Worth, TX CSA
Denver-Aurora, CO CSA
Detroit-Warren-Ann Arbor, MI CSA
Grand Rapids-Wyoming-Muskegon, MI CSA
Hartford-West Hartford, CT CSA
Houston-The Woodlands, TX CSA
Los Angeles-Long Beach, CA CSA
Miami-Fort Lauderdale-Port St. Lucie, FL CSA
Milwaukee-Racine-Waukesha, WI CSA
Minneapolis-St. Paul, MN-WI CSA
New York-Newark, NY-NJ-CT-PA CSA
Philadelphia-Reading-Camden, PA-NJ-DE-MD CSA
Phoenix-Mesa-Scottsdale, AZ MSA
Pittsburgh-New Castle-Weirton, PA-OH-WV CSA
Portland-Vancouver-Salem, OR-WA MSA
Raleigh-Durham-Chapel Hill, NC CSA
Salt Lake City-Provo-Orem, UT CSA
San Diego-Carlsbad, CA MSA
San Jose-San Francisco-Oakland, CA CSA
Seattle-Tacoma-Olympia, WA CSA
Washington-Baltimore-Arlington, DC-MD-VA-WV CSA
R&D Density (R&D spend / km²)
R&D Density (R&D spend / population)

These graphs indicate how dense the innovative inputs (R&D spending) and outputs (patent count) are in some of the largest CSAs and MSAs across the country. If the data was perfectly correlated then each bar in the chart would be at a consistent height. However, since there is tremendous variation specific areas such as Austin, San Diego, and San Jose, for the patent data, and areas such as Detroit, San Jose, and Seattle, for the R&D data indicate that the overall trend is weak. This could be due to a plethora of variables, however it is likely that each city has its own set of reasons. Take for example the high R&D density for Seattle. With the large presence of major corporations such as Amazon, Microsoft, and Boeing the city spends a disproportionate amount of money on R&D. Population and population density do not necessarily indicate why these major firms began and have stayed in the greater Seattle region.

The second and third major conclusions of this study are tougher to represent graphically or with tables, however they are seemingly valid and noteworthy conclusions.

V. CONCLUSION

This paper ventured into new territory by looking at innovation through a completely new lens in the context of population. While the total population of an area has been utilized in previous studies to examine the marginal benefits gained through agglomeration, population density has only been examined in a few different instances. What makes this paper unique is that population density is observed at a CSA/MSA level. In addition the inputs of innovation were also considered by examining R&D data.

While ultimately no strong correlations were found between innovation and population density, the correlation was positive and seems to align with previous research that has been conducted on the topic. In terms of completing further research on the topic it might be worth considering agglomeration at a very granular level. A couple studies have looked at innovation and population density at a zip code level, however if it were possible to drill the data down even further we might be able to get a better glimpse of the true advantages and disadvantages of agglomeration and use of urban space. Perhaps there is even a specific level of density when the marginal costs outweigh the marginal benefits, and an increase in density leads to a decrease in innovation.

Ultimately this research paper illustrates the positive but weak relationship between population density and innovation. In general this conclusion seems to be supported by other historical papers, despite vast economic, social, and political evolutions and changes over time.

VI. REFERENCES CITED

Akcigit, Ufuk, and William R. Kerr. (2010) "Growth Through Heterogeneous Innovations." NBER Working Paper No. 16443.

Berkes, Enrico and Ruben Gaetani (2015) "The Geogrpahy of Unconventional Innovation" 2015 meeting papers 896, Society for Economic Dynamics

Crescenzi, Riccardo, and Andres Rodríguez-Pose, (2013), "R&D, Socio-Economic Conditions, and Regional Innovation in the U.S" Growth Change 44: 287–320

Crescenzi, Riccardo, Andres Rodriguez-Pose and Michael Storper (2007) "The territorial dynamics of innovation: a Europe-United States comparative analysis" Journal of Economic Geography 7: 673-709

Duranton, Gilles, and Diego Puga. 2001 "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products." *American Economic Association*. 91(5): 1454-1477

Florida, Richard, and Charlotta Mellander. (2016) "Rise of the Startup City." *California Management Review* 59 (1): 14-38.

Forman, Chris, Avi Goldfarb and Shane Greenstein. (2016) "Agglomeration of Invention in the Bay Area: Not Just ICT." *American Economic Review*, 106(5): 146-51.

National Science Foundation. 2016. Business Research and Development and Innovation: 2013. Detailed Statistical Tables NSF 16-313. Arlington, VA. Available at http://www.nsf.gov/statistics/srvyindustry/

OECD/Eurostat (2005), Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data, 3rd Edition, *OECD Publishing, Paris*.

OhUallicháin, Breandán. (2002) “Patent Places: Size Matters.” *Journal of Regional Science*, 39(4): 613-636.

Packalen, Mikko, and Jay Bhattacharya. (2015) "Cities and Ideas." NBER. The National Bureau of Economic Research, Jan. 2015. Web. 29 Mar. 2017.

Shambaugh, Jay, Ryan Nunn, and Becca Portman (2017) *Eleven Facts About Innovation and Patents*. Washington, DC: Brookings Institution/Hamilton Project

Thompson Reuters Derwent 2017 World Patents Index on STN database summary Available at https://www.thomsonreuters.com/en/products-services/intellectual-property/patent-research-and-analysis/derwent-world-patents-index.html

US Census Bureau (2017) "Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2016 - United States -- Combined Statistical Area; and for Puerto Rico". (2017) Population Division.